APPENDIX A

Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 1129–1133, February 1988
Evolution

Possible evolution of splice-junction signals in eukaryotic genes from stop codons

(intron elimination/selective pressure/stop-codon walk scanning mechanism/statistical analysis/splicing machinery)

PERIANNAN SENAPATHY*

Division of Computer Research and Technology, National Institutes of Health, Bethesda, MD 20892

Communicated by V. Ramalingaswami, October 21, 1987

ABSTRACT **Splice-junction sequence signals are strongly conserved structural components of eukaryotic genes. These sequences border exon/intron junctions and aid in the process of removing introns by the RNA splicing machinery. Although substantial research has been undertaken to understand the mechanism of splicing, little is known about the origin and evolution of these splice signal sequences. Based on the previously published theory that the primitive genes evolved in pieces from primordial genetic sequences to avoid the interfering stop codons, a "stop-codon walk" mechanism is proposed in this paper to have assisted in the evolution of coding genes. This mechanism predicts the presence of stop codons in splice-junction signals inside the introns. Evidence of the consistent presence of stop codons in the splice-junction signals, in a position where they are expected, is shown by the analysis of codon statistics in these signal sequences in the GenBank databank. The results suggest that the splice-junction signals may have evolved from stop codons as a consequence of a selective pressure to avoid stop codons during the original evolution of coding genes. They also suggest that other splice signals within the introns, such as the branch-point sequence, may have evolved from stop codons for similar reasons.**

The architectural details of a typical eukaryotic gene with its separation into coding sequences (exons) and intervening sequences (introns) remains a matter of intense investigation a decade after its discovery. The mechanism of splicing, by which introns are eliminated and exons are connected up, is becoming understood in increasing detail (for a review, see refs. 1 and 2). It has been found that sequences immediately bordering splice junctions are strongly conserved in genes of a wide variety of eukaryotic organisms, ranging from yeast to humans (3). A sequence of 9 nucleotides is highly conserved at the boundary between an exon and an intron, the donor site. The boundary between an intron and an exon, the acceptor site, also exhibits a highly conserved sequence of 4 nucleotides, preceded by a pyrimidine-rich region. These short conserved sequences are an essential part of the process of exon splicing and provide a specific molecular signal by which the RNA splicing machinery can identify the precise splice points. Although substantial research has been undertaken on the role of splice signals in the RNA splicing process, little is known about the biological meaning of these signals and the mechanism by which they originated. The evolutionary history of these signals could unravel important details about their present-day functions in RNA splicing.

In this paper, I will examine the splice-junction signals for codon statistics and show that they contain a memory of their origin and development that agrees with the following theory of intron evolution that I have presented in detail elsewhere (2). The key finding was that a random sequence of >200 contiguous codons without a stop codon was highly improbable. This same length of 200 codons was also the approximate upper limit of observed exon length; indeed, the theoretical probability distribution for shorter coding sequences also agreed remarkably with the distribution found in actual eukaryotic DNA sequences. The reading-frame lengths were distributed in a negative exponential manner in eukaryotic DNA sequences and computer-generated random DNA sequences. The shortest reading-frame length (zero) was the most frequent. At increasing lengths, the frequency of reading frames decreased negative exponentially, reaching zero at lengths of >600 nucleotides. Thus, it can be supposed that the coding-sequence pieces in primordial DNA or RNA were not longer than modern exons. However, in the first primitive cells, the main selective pressure in evolving a coding gene must have been to generate long coding sequences from the short coding sequences (which existed with an upper limit of 600 nucleotides in the primordial sequences). The long genes of primitive unicellular eukaryotes must have arisen when a mechanism to connect some of the larger primitive coding pieces developed. This would work by splicing out intervening sequences that contained not only the limiting stop codon but also stop codons so closely clustered as to leave only small and inconsequential coding regions.

If the splicing mechanism is an adaptation for eliminating stop codons, it would seem that the splice-junction signals must have evolved in close association with them. The purpose of this paper is to show that this is indeed true and that fossilized stop codons and parts of stop codons are still imbedded in splice-junction signals of present-day eukaryotic genes. The analyses also indicate that the intron elimination by the RNA processing machinery seems to have been mainly geared toward the elimination of stop codons from the primary RNA and that the other splicing signals and possibly components of the splicing machinery may have evolved for this purpose.

RATIONALE

There probably existed strong constraints in the length of coding pieces in primordial genetic sequences due to randomness of nucleotide distribution (2). To avoid the problem of stop-codon interference, the very first coding genes must have evolved with pieces of coding sequences (exons) interrupted by noncoding (intron) sequences. Thus, introns evolved primarily as a means of avoiding the interfering stop codons and increasing the length of the reading frames, resulting in longer polypeptides.

The primary idea here is that a mechanism must have searched for stop codons to define the coding and the noncoding sequences. This "stop-codon walk" mechanism

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*Present address: Biotechnology Center, 1710 University Avenue, University of Wisconsin, Madison, WI 53706.

must have defined the end of an exon when it encountered a stop codon, marking it as the beginning of the intron. This process would have resulted in the evolution of the splice-junction signals containing the stop codons.

RESULTS AND DISCUSSION

A Stop-Codon Walk Mechanism for Coding-Sequence Evolution. Recently I advanced a hypothesis that stop codons played a crucial role in the evolution of the primitive protein-coding genes, which explains why and how the introns first originated (2). The theory was derived as follows. Computer simulations demonstrated that the reading-frame lengths in a random nucleotide sequence are distributed in a negative exponential manner and that there exists an upper limit of about 200 codons in the length of reading frames. Thus, if primordial DNA or RNA contained random nucleotide sequences, there would have been a selective pressure to eliminate interfering stop codons to increase the length of reading frames in evolving a coding gene. Because of the clustering of stop codons, characteristic of a random sequence, single point mutations would not have produced reading frames significantly longer than 600 nucleotides. Therefore, it was suggested that the only way a coding sequence of >600 nucleotides could have ariseh was through a mechanism capable of choosing pieces of coding sequences from within the short reading frames and skipping fairly long sequence stretches to avoid the interfering stop codons (see Fig. 1*A*).

This theory predicts that there must have existed some mechanism to scan a sequence (starting from the initiator codon) to identify a region with interfering stop codons and mark it as an intron. In other words, a stop-codon walk scanning mechanism (somewhat analogous to the reading of mRNA by ribosomes) could have operated in the evolution of protein-coding genes (see Fig. 1*B*). This mechanism, reading downstream from the initiator codon, must have looked for and eliminated the first occurring stop codon by marking that stop codon as the beginning of an intron. After skipping that intron stretch, which usually contains a cluster of stop codons, and then finding a fairly large piece of coding sequence, the machinery must have again looked for the first interfering stop codon and marked it as the beginning of the next intron. This search mechanism for stop codons, which was subject to natural selection over a long period of time, must have evolved the junction signals for splicing. The stop

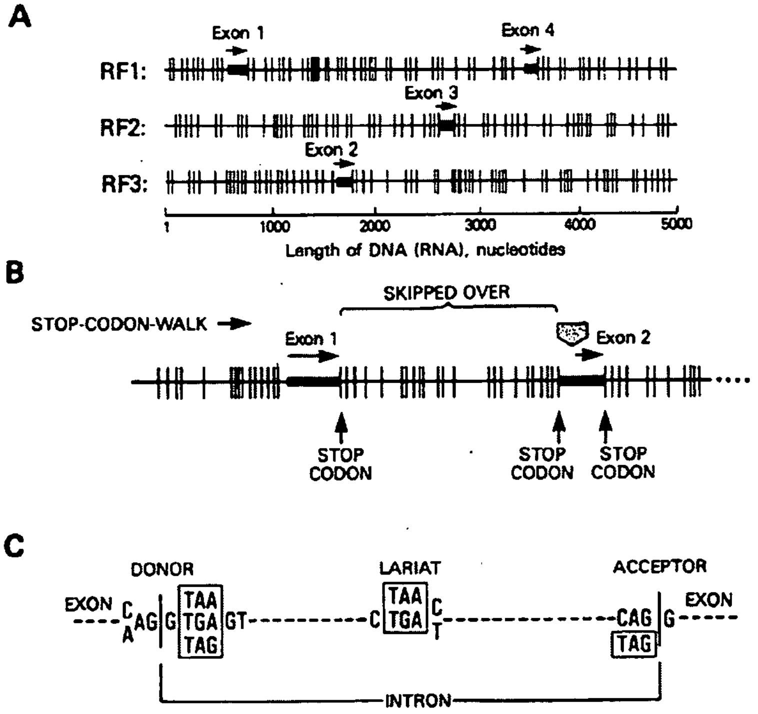

FIG. 1. Evolution of introns to eliminate interfering stop codons, making stop codons part of splice-junction signals. (*A*) In a simulated random sequence of nucleotides, the stop codons occurred in a fashion restricting the length of reading frames to about 600 nucleotides (even in a sequence of the order of billions of nucleotides). Thus, original coding genes, which presumably evolved from random primordial genetic sequences, were forced to evolve in short pieces (thick lines), skipping stretches of sequences to avoid the clusters of stop codons (tick marks). The pieces of coding sequences were represented in all three reading frames, RF1, RF2, and RF3. The spliced coding pieces then contiguously coded for a protein. (*B*) A stop-codon walk mechanism (that defined the coding and intervening sequences in a sequential manner) probably identified the first occurring stop codon as the beginning of the intervening sequence. By natural selection, the first stop codon became an integral part of the splice-junction signal such that the stop codon was eliminated along with the intron. A stop codon may or may not have preceded the beginning of an exon (broad arrow) conferring less selective pressure to eliminate a stop codon at the beginning of an exon. (*C*) Two of the stop codons occur with very high frequency at the start of today's introns (TAA and TGA far more more frequently than TAG) and only one occurs at their ends (TAG). A conserved sequence toward the end of introns (about 30 nucleotides upstream from the 3′ end) that forms a lariat structure with the beginning of the intron also consistently contains stop codons. Both the A-G of the ${}^{\mathrm{C}}_{\mathrm{T}}$AG at the end of introns and the ${}^{\mathrm{C}}_{\mathrm{A}}$AG at the beginning of introns seem to have come from the original selective pressure to eliminate TAG at the end of introns. The presence of different stop codons in the donor and acceptor signals may indicate that this difference evolved for the splicing machinery to precisely differentiate the two ends of an intron. Only the stop codons (TAA, TAG, TGA) are boxed in the figure.

Donor Splice Signal	$^{C}_{A}$AG:GT$^{A}_{G}$AGT
Acceptor Splice Signal	$^{C}_{T}$AG:G
Lariat Signal	CT$^{A}_{G}$A$^{C}_{T}$
Poly-A Addition Signal	AATAAA

FIG. 2. Consensus sequences for the donor splice signal, acceptor splice signal, and the lariat signal are shown. The colon indicates the splice point. The donor signal occurs at the exon/intron junction and the acceptor signal occurs at the intron/exon junction. The lariat signal occurs at about 20–50 nucleotides upstream of the acceptor signal in the intron.

codons would thus have become a part of the splice-junction signals and in such a position as to be eliminated during splicing along with the introns (see Fig. 1*B*).

The Consistent Presence of a Stop Codon in the Splice-Junction Signal Inside the Intron. Breathnach *et al.* (3) first recognized that sequences immediately surrounding the splice points are highly conserved, and that a "consensus" sequence could be arrived at by tabulating the nucleotide frequencies in these regions around a large number of splice points (3, 4). Recently, consensus sequences have been built from larger sequence data (5). For the exon/intron (donor) junction, the consensus sequence is $^{C}_{A}$AG:GT$^{A}_{G}$AGT; for the intron/exon (acceptor) junction, it is $^{C}_{T}$AG:G (Fig. 2). These splice junction sequences seem to serve as guides for the "spliceosome" machinery to identify the precise splice points (1).

The codon statistics in splice-junction sequences of today's eukaryotic genes could reveal whether a stop-codon walk mechanism did in fact play a role in the evolution of coding genes. Examination of the codon frequencies at the seven possible nucleotide positions around the donor junction shows that all three stop codons occur with very high frequency on the intron side, at the fifth nucleotide position, 1 nucleotide immediately downstream of the splice point. (See Table 1; out of a total of 1030 donor sites in the GenBank databank,† the stop codons occur in 726 at the fifth nucleotide position.) The expected mean frequency of stop codons at any nucleotide position is 3/64, which is 55 out of 1030 sites. The donor consensus sequence $^{C}_{A}$AG:GT$^{A}_{G}$AGT has all its nucleotide positions invariant downstream of the splice point, except the adenine or guanine variance at the sixth nucleotide position, leading precisely to the TAA, TGA, or TAG codon. Thus, starting at the fifth position, there is a very high frequency of stop codons. My theory of the stop-codon walk scanning mechanism would predict a stop codon immediately after the donor splice point, and, in fact, it starts 1 nucleotide further downstream. These facts support my view that the splice-junction signals evolved from stop codons. This possibility is further strengthened by the fact that most of the other codons beginning at the fifth nucleotide position (see Table 1) start with T·A or T·G, the first 2 nucleotides of the three stop codons.

Strong Donor, Weak Acceptor Stop-Codon Pattern. Thus, this theory predicts the presence of stop codons in the donor splice-junction signal and explains the meaning of this signal sequence in a consistent manner. It also suggests that the frequency of stop codons at the acceptor signal must be low, because the pressure to eliminate stop codons at the start of exons would not have been so pronounced as at the end of exons. The mechanism had to eliminate a stop codon at the front end of an exon only if it encountered a stop codon immediately upstream of the beginning of the exon (Fig. 1*B*). Thus, the selective pressure to eliminate stop codons would have been very strong at the 3′ end of an exon and weak at the 5′ end. Analysis of the acceptor junction sequences (Fig. 2 and Table 1) reveals that stop codons occur with less frequency preceding the beginning of exons (out of 1030 acceptor sites in the GenBank databank, TAG occurs in 234 at the −3 position) than following the end of exons. Thus, there exists a high frequency of stop codons in the donor signal and a low frequency in the acceptor signal, as suggested by the theory (2).

Table 1. Frequency of stop codons in donor and acceptor splice-junction sequences

Codon	Number of occurrences in donor signal	Number of occurrences in acceptor signal
TAA	370	0
TGA	292	0
TAG	64	234
CAG	7	746
Other	297*	50
Total	1030	1030

A computer program was used to count codons at specified nucleotide positions relative to the splice points. Codon counts were made from each of all of the possible 7-nucleotide positions in the donor signal [$^{C}_{A}$AG:GT$^{A}_{G}$AGT] and each of the possible 2-nucleotide positions in the acceptor signal [$^{C}_{T}$AG:G] from all of the protein-coding gene sequences from the GenBank databank. High frequency of stop codons was found only at the fifth nucleotide position in the donor signal and the first nucleotide position in the acceptor signal. The codon counts at only these positions are shown. Some relevant codon counts at the first nucleotide position in the donor signal are CAG = 244; AAG = 165; GAG = 76; TAG = 15; TGA = 8; TAA = 4.

*More than 70% are T$^{G}_{A}$X [TAT = 75; TAC = 59; TG$\overset{C}{\underset{G}{T}}$ = 70].

Stop Codons in Splice-Junction Signals of Only Protein-Coding Genes. The protein-coding genes in organisms as diverse as yeasts, insects, chickens, and humans exhibit consensus splice-junction signal sequences containing stop codons. The rRNA and tRNA genes from these organisms (4), although they contain introns and use splicing, do not have splicing signals based on stop codons. The functions of rRNA and tRNA genes (which do not code for proteins) are not expected to be affected by the presence of stop codons in their sequences. The consistent presence of stop codons in the splice junctions of only the protein-coding genes seems to be a strong reason to believe that stop codons have played a role in the origin and evolution of splice-junction signal sequences. Splicing in tRNAs and rRNAs might have been necessary for transportation from the nucleus to the cytoplasm and may have been evolved for that reason.

Other Remnants of the Stop-Codon Walk Machinery. A short stretch of conserved nucleotides present in introns 20–50 nucleotides upstream of the acceptor splice junction has been shown to aid the splicing process by first creating a "lariat" structure with the preceding donor site (6, 7). The consensus sequence at the lariat site for yeast is TACTAAC (8, 9). Genes of higher eukaryotes also contain a potential lariat consensus sequence, CT$^{G}_{A}$A$^{C}_{T}$ (Fig. 2) (10, 11). In both, the presence of the stop codons is clearly evident, TGA or TAA being consistently found within a short sequence of 5 nucleotides (Fig. 1*C*). The branching point of the lariat

†EMBL/GenBank Genetic Sequence Database (1987) GenBank (Bolt, Beranek, and Newman Laboratories, Cambridge, MA), Tape Release 46.0.

occurs at the last adenine of the stop codon. This leads to the possibility that the machinery for the elimination of stop codons originally created an auxiliary stop-codon sequence signal (the lariat sequence) to aid its splicing function.

The small nuclear U2 RNA found in splicing complexes is thought to aid splicing by interacting with the lariat sequence (11). Complementary sequences for both the lariat sequence and the acceptor signal are present in a segment of only 15 nucleotides in U2 RNA. Further, the U1 RNA has been proposed to function as a guide in splicing to identify the precise donor splice junction by complementary base-pairing. The conserved regions of the U1 RNA thus include sequences complementary to the stop codons. These observations, in light of the present findings, may indicate that stop codons had operated in the evolution of not only the splice-junction signals and the lariat signal but also of some of the small nuclear RNAs.

A conserved sequence, AATAAA, exists in almost every gene a short distance downstream from the end of the protein-coding message and serves as a signal for the addition of poly(A) in the mRNA copy of the gene (Fig. 2) (12). This poly(A) sequence signal contains a stop codon, TAA. A sequence shortly downstream from this signal (13), thought to be part of the complete poly(A) signal, also contains the TAG and TGA stop codons. Thus, the evolution of the whole RNA processing mechanism seems to have been geared toward elimination of stop codons, thus making those stop codons the focal points for RNA processing.

Different Stop Codons in the Donor and Acceptor Junction Signals. The theory of stop-codon elimination (2) would predict the presence of all three stop codons in both the donor and acceptor junction signals. However, the codon statistics for the donor and acceptor signals show that they contain different stop codons. In the donor signal, the most frequent stop codons are TAA and TGA, whereas TAG occurs rarely. In the acceptor signal, only TAG occurs, with absolutely no TAA and TGA (Fig. 1*C* and Table 1). This clear difference may have evolved because the two ends of the intron had to be marked differently, for the splicing mechanism to distinguish between the donor and acceptor sites.

Although the mechanism for eliminating stop codons may have evolved a cooperative recognition between the donor and acceptor junction signals, with the main components being stop codons, apparently, sequences other than stop codons also have been used in the recognition process. Even the evolution of these other nucleotide positions seems to have been influenced by the selection pressure to eliminate stop codons. This is seen from the codon statistics around splice-junction sequences. As seen above, the donor contains the two stop codons TAA and TGA, and the acceptor contains only TAG. At the start of the acceptor signal, virtually the only codon found other than TAG is CAG. (The codon count, at the −3 position, for TAG is 234 and for CAG is 746 out of 1030 acceptor sites; Table 1.) Not a single TGA or TAA is present at this position. In contrast, the start of the donor signal and the start of the acceptor signal are quite similar; in fact, the acceptor signal is almost a subset of the donor signal. Starting with the theme that stop codons were crucial, the data indicate that the original acceptor sequence must have had TAG. Thus, the fact that the present-day acceptor signal more often starts with CAG suggests that the acceptor signal CAG was established because of its similarity to TAG (that is, the A·G may have originated from TAG). There must have been some mechanistic need in the evolving splicing machinery for such an A·G sequence to be also at the start of the donor splice signal. This mechanistic need for a similarity in the donor and acceptor starting sequences must have caused the evolution of $^{C}_{A}$AG at the donor first codon position, keeping the A·G that originated from TAG of the acceptor.

Reading Frame of the Stop Codons in the Splice Signal Sequences. The hypothesis for the evolution of split genes (2) suggests that the process of splicing evolved in the primordial soup primarily to increase the length of reading frames without a predetermined purpose for any resulting protein. Each newly spliced RNA may or may not have coded for a protein that was functional. At this early stage of evolution the process of splicing may have led to the synthesis of numerous long polypeptides, most having no biological function. It would have taken further natural selection to identify those rare proteins with advantageous functions. Thus, a functional exon becomes the afterfact of the natural selection of a polypeptide having a selective advantage.

Questions arise as to why the introns should be long in general and why some introns, containing open reading frames as long as 600 nucleotides, are not used by the particular gene in question as part of its coding sequence. To address these questions one must consider the characteristics necessary for a viable splice junction. The stop-codon scanning mechanism would have selected exons and introns based on (*i*) the absence of interfering stop codons in the final spliced exons, (*ii*) the availability of the splice-junction signals precisely at the ends of the exons, (*iii*) the consecutive sequential combinations of splice signals and branch-point signals, and (*iv*) the absence of these signals at the wrong places. All of these conditions can be satisfied only within a sequence of considerable length and hence long introns may have been the natural outcome. Because of these reasons it is possible that two genes may overlap one another, and the intron of one gene can contain an exon of another gene.

The subject of stop-codon reading frames in the splice signals deserves mention here. My theory suggests that stop codons in the splice signals, which were sought out to be eliminated in evolution, must be in the same reading frame as that of the original open reading frame in the evolved genes. However, once the splice signals were established by the stop-codon scanning mechanism in evolution, later, the presence of a strong splice-junction signal even before the occurrence of an interfering stop codon in a sequence may indicate a splicing point, with the primary requirement that the final spliced exons had no interfering stop codons. Thus, once the splicing machinery had evolved to recognize a splice signal, the actual location of the stop codon within it would not be important nor would its reading frame relative to the previous exon, since the stop codon sequence itself would be a part of the splice signal on the intron side that is spliced out.

The stop codons in the splice junction signals can exist in any of the three reading frames because the splice junctions are not at the end of a codon triplet and because the lengths of exons are not complete codon lengths. Only if the splice junctions were at the end of a codon would the reading frames of consecutive exons remain unchanged. Further, the intron length is also not constrained to a multiple of three, so the reading frame does not hold through the intron. Indeed, evolving a coding gene from pieces of sequences in all the three reading frames is more advantageous than evolving from sequences in the same reading frame (unpublished).

Possible Recognition of Stop Codons in Today's RNA Splicing Machinery. Thus, when the splicing mechanism originally evolved to avoid the too-frequent occurrences of stop codons in primordial DNA or RNA, it may have had several ways of recognizing stop codons and perhaps used complementary RNA guide sequences. While across evolutionary time natural selection may have significantly altered splicing mechanisms, they may very well still retain some original features of the splicing machinery, including features of the stop-codon walk process. Thus, the small nuclear RNAs, probably used as guides to recognize stop-codon signals by the primordial splicing machinery, may still be used for the same purpose in today's RNA splicing process. Therefore, it will not be surprising if some of the small nuclear RNAs associated with spliceosomes (1) are found to be used for recognizing stop-codon signals. Further, originally the splicing machinery itself may have searched for stop codons, identified exons and introns, and also carried out the splicing process. However, it is possible that the functions of searching for stop codons and identifying introns have been partly lost from the splicing machinery, leaving only the splicing mechanism. If, on the other hand, some components of today's spliceosomes are used for recognizing stop codons, it must be possible to verify this phenomenon by binding experiments.

Various researchers have analyzed the importance of mutational changes in the different nucleotide positions of splice signals (14, 15) and lariat signal (10) in the splicing activity. Some of these mutational changes decrease or eliminate the production of accurately spliced RNA *in vivo*, whereas others have little effect. These studies have shown that the G·T (fourth and fifth positions) in the donor, the A·G in the acceptor, and the last adenine in the lariat signal are the most important in splicing. These are, with the exception of the guanine in the G·T, parts of stop codons. More detailed studies of this kind can be expected to show more clearly that stop codons as well as parts of stop codons are crucial in splicing.

Thus, the possible evolution of splice-junction signals from stop-codon sequences by a stop-codon walk mechanism seems to be supported in three respects: (*i*) the concurrence of the upper limit of exon length in actual eukaryotic DNA sequences with the upper limit of reading-frame length in a random sequence (2); (*ii*) the striking similarity between reading-frame length statistics in actual eukaryotic DNA sequences and random sequences; and (*iii*) the consistent presence of stop codons in the splice-junction signals. In conclusion, the statistical information from published gene sequences indicates that the splice-junction signals may have evolved from stop-codon sequences because of selective pressure to avoid stop codons during the evolution of the first coding genes. The "memory" of this phenomenon still lingers in splice-junction sequences of present-day eukaryotic genes and may still serve crucial functions in the process of splicing in today's eukaryotes.

I thank Drs. R. Shaw, R. Mage, and E. Ackerman and Mr. S. Knisley and Ms. M. Hodges for help in improving the manuscript and Mr. M. Shapiro for writing the computer program for counting codons. I am grateful to Dr. Arnold W. Pratt for his constant encouragement.

1. Sharp, P. A. (1987) *Science* **235**, 766–771.
2. Senapathy, P. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 2133–2137.
3. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. & Chambon, P. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 4853–4857.
4. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50**, 349–383.
5. Shapiro, M. B. & Senapathy, P. (1987) *Nucleic Acids Res.* **15**, 7155–7174.
6. Grabowski, P. J., Padgett, R. A. & Sharp, P. A. (1984) *Cell* **137**, 415–427.
7. Ruskin, B., Krainer, A. R., Maniatis, T. & Green, M. R. (1984) *Cell* **38**, 317–331.
8. Langfold, C. J., Klinz, F. J., Donath, C. & Gallwitz, D. (1984) *Cell* **36**, 645–653.
9. Piekielny, C. W., Teem, J. L. & Rosbash, M. (1981) *Cell* **34**, 395–403.
10. Rautmann, G. & Breathnach, R. (1985) *Nature (London)* **315**, 430–432.
11. Keller, E. B. & Noon, W. A. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 7417–7420.
12. Birnstiel, M. L., Busslinger, M. & Strub, K. (1985) *Cell* **41**, 349–359.
13. McDevitt, M. A., Imperiate, M., Ali, H. & Nevins, J. R. (1984) *Cell* **37**, 993–999.
14. Wieringa, B., Hofer, E. & Weissmann, C. (1984) *Cell* **37**, 915–925.
15. Ruskin, B. & Green, M. R. (1985) *Nature (London)* **317**, 732–734.